EXHIBIT 4

SENATE JUDICIARY COMMITTEE
Senator Thomas Umberg, Chair
2023-2024 Regular Session

AB 2655 (Berman)
Version: June 11, 2024
Hearing Date: July 2, 2024
Fiscal: Yes
Urgency: No
CK

SUBJECT

Defending Democracy from Deepfake Deception Act of 2024

DIGEST

This bill establishes the Defending Democracy from Deepfake Deception Act of 2024, which requires a large online platform to block the posting or sending of materially deceptive and digitally modified or created content related to elections, during specified periods before and after an election. It requires these platforms to label certain additional content inauthentic, fake, or false during specified periods before and after an election and to provide mechanisms to report such content.

EXECUTIVE SUMMARY

The rapid advancement of AI technology, specifically the wide-scale introduction of generative AI models, has made it drastically cheaper and easier to produce synthetic content – audio, images, text, and video recordings that are not real, but that are so realistic that they are virtually impossible to distinguish from authentic content, including so-called "deepfakes." In the context of election campaigns, such deepfakes can be weaponized to deceive voters into thinking that a candidate said or did something which the candidate did not, or otherwise falsely call election results into question. A series of bills currently pending before this Committee attempt to address these issues by restricting or labeling AI-altered or –generated content. However, this bill specifically targets social media platforms and such materially deceptive content on their platforms, requiring platforms to block and prevent it, label it, and provide mechanisms for reporting it.

The bill is sponsored by the California Initiative for Technology & Democracy. It is supported by various organizations, including the League of Women Voters of California and Disability Rights California. It is opposed by Oakland Privacy and various industry associations, including TechNet. The bill passed out of the Senate Elections and Constitutional Amendments Committee on a 6 to 1 vote.

PROPOSED CHANGES TO THE LAW

Existing law:

1) Provides that "Congress shall make no law… abridging the freedom of speech..." (U.S. Const., amend. 1.)

2) Applies the First Amendment to the states through operation of the Fourteenth Amendment. (*Gitlow v. New York* (1925) 268 U.S. 652; *NAACP v. Alabama* (1925) 357 U.S. 449.)

3) Provides, in federal law, that a provider or user of an interactive computer service shall not be treated as the publisher or speaker of any information provided by another information content provider. (47 U.S.C. § 230(c)(1).)

4) Provides that a provider or user of an interactive computer service shall not be held liable on account of:
   a. any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or
   b. any action taken to enable or make available to information content providers or others the technical means to restrict access to such material. (47 U.S.C. § 230(c)(2).)

5) Provides that no provider or user of an interactive computer service shall be treated for liability purposes as the publisher or speaker of any information provided by another information content provider. (47 U.S.C. § 230.)

6) Defines "materially deceptive audio or visual media" as an image or an audio or video recording of a candidate's appearance, speech, or conduct that has been intentionally manipulated in a manner such that both of the following conditions are met:
   a. The image or audio or video recording would falsely appear to a reasonable person to be authentic.
   b. The image or audio or video recording would cause a reasonable person to have a fundamentally different understanding or impression of the expressive content of the image or audio or video recording than that person would have if the person were hearing or seeing the unaltered, original version of the image or audio or video recording. (Elec. Code § 20010(e).)

7) Prohibits a person, committee, or other entity from distributing with actual malice materially deceptive audio or visual media of a candidate with the intent to injure the candidate's reputation or to deceive a voter into voting for or against the candidate within 60 days of an election at which a candidate for elective office will appear on the ballot, as specified and unless specified conditions are met. (Elec. Code § 20010(a).)

8) Exempts audio or visual media that includes a disclosure stating: "This _____ has been manipulated." Requires the blank in the disclosure to be filled with a term that most accurately describes the media, as specified. Requires the following disclosures for visual and audio-only media:
   a. For visual media, the text of the disclosure shall appear in a size that is easily readable by the average viewer and no smaller than the largest font size of other text appearing in the visual media. If the visual media does not include any other text, then the disclosure shall appear in a size that is easily readable by the average viewer. Requires, for visual media that is video, the disclosure to be displayed throughout the duration of the video.
   b. For audio-only media, the disclosure shall be read in the clearly spoken manner and in a pitch that can be easily heard by the average listener, at the beginning of the audio, at the end of the audio, and, if the audio is greater than two minutes in length, interspersed within the audio at intervals of not greater than two minutes each. (Elec. Code § 20010(b).)

9) Permits a candidate for elective office whose voice or likeness appears in a materially deceptive audio or visual media distributed in violation of the above provisions, to seek injunctive or other equitable relief prohibiting the distribution of audio or visual media in violation. (Elec. Code § 20010(c)(1).)

10) Permits a candidate for elective office whose voice or likeness appears in materially deceptive audio or visual media distributed in violation of the above provisions to bring an action for general or special damages against the person, committee, or other entity that distributed the materially deceptive audio or visual media, as specified. Requires the plaintiff to bear the burden of establishing the violation through clear and convincing evidence in any civil action alleging a violation, as specified. (Elec. Code § 21101(c)(2).)

This bill:

1) Establishes the Defending Democracy from Deepfake Deception Act of 2024.

2) Requires a large online platform, using state-of-the-art, best available tools to detect materially deceptive content, to develop and implement procedures for blocking and preventing, and, if the platform knows or should know that the materially deceptive content meets the requirements hereof, to block and prevent

the posting or sending of any materially deceptive content, if all of the following conditions are met:

a) The content is posted or sent during a period beginning 120 days before the election and through the day of the election. For content that depicts or pertains to elections officials, this period shall extend to the 60th day after the election.
b) The materially deceptive content is any of the following:
    i. A candidate for elective office portrayed as doing or saying something that the candidate did not do or say and that is reasonably likely to harm the reputation or electoral prospects of a candidate.
    ii. An elections official portrayed as doing or saying something in connection with the performance of their elections-related duties that the elections official did not do or say and that is reasonably likely to falsely undermine confidence in the outcome of one or more election contests.
    iii. An elected official portrayed as doing or saying something that influences the election that the elected official did not do or say and that is reasonably likely to falsely undermine confidence in the outcome of one or more election contests.
c) The person or entity who created the materially deceptive content did so knowing it was false or with reckless disregard for the truth. There shall be a rebuttable presumption that the person or entity knew the materially deceptive content was false or acted with reckless disregard for the truth if the content would cause a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if hearing or seeing an authentic version of the content.

3) Requires, notwithstanding the above, a large online platform to allow a candidate for elective office, during a period beginning 120 days before the election and through the day of the election, to portray themself as doing or saying something that the candidate did not do or say, but only if the digital content includes a disclosure meeting specified conditions and states the following: "This [category of content] has been manipulated."

4) Requires a large online platform, using state-of-the-art, best available tools to detect materially deceptive content to develop and implement procedures for labeling such content as inauthentic, fake, or false if all of the following conditions are met:
    a) The materially deceptive content is either of the following:
        i. Meets the standards set above, but is posted or sent outside the applicable time period.
        ii. Appears within an advertisement or election communication and is not subject to the above.

b) The person or entity who created the materially deceptive content did so knowing it was false or with reckless disregard for the truth. There shall be a rebuttable presumption that the person or entity knew the materially deceptive content was false or acted with reckless disregard for the truth if the content would cause a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if hearing or seeing an authentic version of the content.
c) The large online platform knows or should know that the materially deceptive content meets the requirements of this section.

5) Specifies required functionality of the label above and states the labeling requirement applies during redistricting and during a period from one year before the election through election day. If the content involves elections officials, the electoral college process, the canvass of the vote, or election-related equipment or property, the time period is extended 60 days beyond the election.

6) Requires a large online platform to provide an easily accessible way for California residents to report to that platform content subject to the above provisions that was not blocked or labeled as required. The online platform shall respond to the person who made the report, within 36 hours of the report, describing any action taken or not taken by the online platform with respect to the content.

7) Authorizes a candidate for elective office, elected official, or elections official who has made a report and who either has not received a response within 36 hours or disagrees with the response, as well as the Attorney General or any district attorney or city attorney, to seek injunctive or other equitable relief against the online platform to compel compliance. The plaintiff shall bear the burden of establishing the violation through clear and convincing evidence. The court is required to award a prevailing plaintiff reasonable attorney's fees and costs. Such actions are given precedence in accordance with Section 35 of the Code of Civil Procedure.

8) Clarifies that it applies to materially deceptive content, regardless of the language used in the content. If the language used is not English, the required disclosure and label must appear in the language used as well as in English.

9) Requires a large online platform that blocks or labels any materially deceptive content to maintain a copy of the digital content for a period of not less than five years from the election and shall make such digital content available to the Secretary of State, the Fair Political Practices Commission, and researchers, if requested.

10) Exempts from the scope of the bill the following:
    a) A regularly published online newspaper, magazine, or other periodical of general circulation that routinely carries news and commentary of general interest, and that publishes any materially deceptive content that an online platform is required to block or label based on this chapter, if the publication contains a clear disclosure that the materially deceptive content does not accurately represent any actual event, occurrence, appearance, speech, or expressive conduct.
    b) Materially deceptive content that constitutes satire or parody.

11) Includes findings and declarations and a severability clause.

12) Defines the relevant terms, including:
    a) "Materially deceptive content" means audio or visual media that is digitally created or modified, and that includes, but is not limited to, deepfakes and chatbots, such that it would falsely appear to a reasonable person to be an authentic record of the content depicted in the media.
    b) "Large online platform" means a public-facing internet website, web application, or digital application, including a social network, video sharing platform, advertising network, or search engine that had at least 1,000,000 California users during the preceding 12 months.

COMMENTS

1. Blurring reality: AI-generated content

Generative AI is a type of artificial intelligence that can create new content, including text, images, code, or music, by learning from existing data. Generative AI models can produce realistic and novel artifacts that resemble the data they were trained on, but do not copy it. For example, generative AI can write a poem, draw a picture, or compose a song based on a given prompt or theme. Generative AI enables users to quickly generate new content based on a variety of inputs. Generative AI models use neural networks to identify the patterns and structures within existing data to generate new and original content.

The world has been in awe of the powers of this generative AI since the widespread introduction of AI systems such as ChatGPT. However, the capabilities of these advanced systems leads to a blurring between reality and fiction. The Brookings Institution lays out the issue:

> Over the last year, generative AI tools have made the jump from research prototype to commercial product. Generative AI models like OpenAI's ChatGPT and Google's Gemini can now generate realistic text and images that are often indistinguishable from human-authored content, with

> generative AI for audio and video not far behind. Given these advances, it's no longer surprising to see AI-generated images of public figures go viral or AI-generated reviews and comments on digital platforms. As such, generative AI models are raising concerns about the credibility of digital content and the ease of producing harmful content going forward.
>
> Against the backdrop of such technological advances, civil society and policymakers have taken increasing interest in ways to distinguish AI-generated content from human-authored content.[1]

One expert at the Copenhagen Institute for Future Studies estimates that should large generative-AI models run amok, up to 99 percent of the internet's content could be AI-generated by 2025 to 2030.[2] The problematic applications are seemingly infinite, whether it be deepfakes to blackmail or shame victims, false impersonations to commit fraud, or other nefarious purposes. Infamously, in January of this year, Taylor Swift was the victim of sexually explicit, nonconsensual deepfake images using AI that were widely spread across social media platforms.[3] Perhaps more disturbingly, a trend has emerged in schools of students creating such images: "At schools across the country, people have used deepfake technology combined with real images of female students to create fraudulent images of nude bodies. The deepfake images can be produced using a cellphone."[4] As more of the population becomes aware of the potential to realistically fake images, video, and text, some will use the skepticism that creates to challenge the authenticity of real content, a phenomena coined the "liar's dividend."[5]

Relevant here, AI and specifically generative AI can spread misinformation regarding elections with ease, both in California and across the world:

> Artificial intelligence is supercharging the threat of election disinformation worldwide, making it easy for anyone with a smartphone

---

[1] Siddarth Srinivasan, *Detecting AI fingerprints: A guide to watermarking and beyond* (January 4, 2024) Brookings Institution, https://www.brookings.edu/articles/detecting-ai-fingerprints-a-guide-to-watermarking-and-beyond/#:~:text=Google%20also%20recently%20announced%20SynthID,model%20to%20detect%20the%20watermark. All internet citations are current as of June 23, 2024.

[2] Lonnie Lee Hood, *Experts Say That Soon, Almost The Entire Internet Could Be Generated by AI* (March 4, 2022) The Byte, https://futurism.com/the-byte/ai-internet-generation.

[3] Brian Contreras, *Tougher AI Policies Could Protect Taylor Swift – And Everyone Else – From Deepfakes* (February 8, 2024) Scientific American, https://www.scientificamerican.com/article/tougher-ai-policies-could-protect-taylor-swift-and-everyone-else-from-deepfakes/.

[4] Hannah Fry, Laguna Beach High School investigates 'inappropriate' AI-generated images of students (April 2, 2024) Los Angeles Times, https://www.latimes.com/california/story/2024-04-02/laguna-beach-high-school-investigating-creation-of-ai-generated-images-of-students.

[5] Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security* (July 14, 2018) 107 California Law Review 1753 (2019), https://ssrn.com/abstract=3213954.

> and a devious imagination to create fake – but convincing – content aimed at fooling voters.
>
> It marks a quantum leap from a few years ago, when creating phony photos, videos or audio clips required teams of people with time, technical skill and money. Now, using free and low-cost generative artificial intelligence services from companies like Google and OpenAI, anyone can create high-quality "deepfakes" with just a simple text prompt.
>
> A wave of AI deepfakes tied to elections in Europe and Asia has coursed through social media for months, serving as a warning for more than 50 countries heading to the polls this year.
>
> "You don't need to look far to see some people ... being clearly confused as to whether something is real or not," said Henry Ajder, a leading expert in generative AI based in Cambridge, England.
>
> The question is no longer whether AI deepfakes could affect elections, but how influential they will be, said Ajder, who runs a consulting firm called Latent Space Advisory.
>
> As the U.S. presidential race heats up, FBI Director Christopher Wray recently warned about the growing threat, saying generative AI makes it easy for "foreign adversaries to engage in malign influence."[6]

On that last note, in February of this year, voters in New Hampshire received robocalls that are purported to have used an AI voice resembling President Joe Biden advising them against voting in the presidential primary and saving their vote for the November general election.[7] The examples are endless:

> Former President Donald Trump, who is running in 2024, has shared AI-generated content with his followers on social media. A manipulated video of CNN host Anderson Cooper that Trump shared on his Truth Social platform on Friday, which distorted Cooper's reaction to the CNN town hall this past week with Trump, was created using an AI voice-cloning tool.

---

[6] Ali Swenson & Kelvin Chan, *Election disinformation takes a big leap with AI being used to deceive worldwide* (March 14, 2024) Associated Press, https://apnews.com/article/artificial-intelligence-elections-disinformation-chatgpt-bc283e7426402f0b4baa7df280a4c3fd.

[7] Em Steck & Andrew Kaczynski, *Fake Joe Biden robocall urges New Hampshire voters not to vote in Tuesday's Democratic primary* (January 22, 2024) CNN, https://www.cnn.com/2024/01/22/politics/fake-joe-biden-robocall/index.html.

> A dystopian campaign ad released last month by the Republican National Committee offers another glimpse of this digitally manipulated future. The online ad, which came after President Joe Biden announced his reelection campaign, and starts with a strange, slightly warped image of Biden and the text "What if the weakest president we've ever had was re-elected?"
>
> A series of AI-generated images follows: Taiwan under attack; boarded up storefronts in the United States as the economy crumbles; soldiers and armored military vehicles patrolling local streets as tattooed criminals and waves of immigrants create panic.
>
> "An AI-generated look into the country's possible future if Joe Biden is re-elected in 2024," reads the ad's description from the RNC.
>
> The RNC acknowledged its use of AI, but others, including nefarious political campaigns and foreign adversaries, will not, said Petko Stoyanov, global chief technology officer at Forcepoint, a cybersecurity company based in Austin, Texas. Stoyanov predicted that groups looking to meddle with U.S. democracy will employ AI and synthetic media as a way to erode trust.[8]

Legislatures across the country are pushing legislation that would address this looming threat.

2. Materially deceptive content in political advertisements

According to the author:

> AB 2655 will ensure that online platforms restrict the spread of election-related deceptive deepfakes meant to prevent voters from voting or to deceive them based on fraudulent content. Deepfakes are a powerful and dangerous tool in the arsenal of those that want to wage disinformation campaigns, and they have the potential to wreak havoc on our democracy by attributing speech and conduct to a person that is false or that never happened. Advances in technology make it easy for practically anyone to generate this deceptive content, making it that much more important that we identify and restrict its spread before it has the chance to deceive voters and undermine our democracy.

---

[8] David Klepper & Ali Swenson, *AI-generated disinformation poses threat of misleading voters in 2024 election* (May 14, 2023) PBS News, https://www.pbs.org/newshour/politics/ai-generated-disinformation-poses-threat-of-misleading-voters-in-2024-election.

Unlike existing law or other bills pending before this Committee in this area, this bill seeks to place responsibility on large online platforms with regard to "materially deceptive content" regarding elections, placing a series of obligations on them. The bill defines "materially deceptive content" as audio or visual media that is digitally created or modified, and that includes, but is not limited to, deepfakes and chatbots, such that it would falsely appear to a reasonable person to be an authentic record of the content depicted in the media. "Large online platform" means a public-facing internet website, web application, or digital application, including a social network, video sharing platform, advertising network, or search engine that had at least 1,000,000 California users during the preceding 12 months.[9]

a. Preventing and blocking materially deceptive content

The bill requires platforms to develop and implement procedures for blocking and preventing, and, to block and prevent the posting or sending of, materially deceptive content, if the platform knows or should know that the materially deceptive content meets the requirements of the bill and certain conditions are met.

The materially deceptive content must portray one of the following. First is content portraying a candidate for elective office as doing or saying something they did not do or say and that is reasonably likely to harm the reputation or electoral prospects of a candidate. Or it must portray an elected official as doing or saying something that influences the election or an elections official as doing or saying something in connection with the performance of their elections-related duties that the official did not do or say and that is reasonably likely to falsely undermine confidence in the outcome of one or more election contests.

Second, the content must be posted or sent during a period beginning 120 days before the election and through the day of the election. For content that depicts or pertains to elections officials, this period shall extend to the 60th day after the election.

Finally, to trigger the requirement for platforms to block and prevent the content, the person or entity who created the content must have done so knowing it was false or with reckless disregard for the truth.

In any ensuing litigation, the bill establishes a rebuttable presumption that the person or entity knew the materially deceptive content was false or acted with reckless disregard for the truth if the content would cause a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if hearing or seeing an authentic version of the content.

---

[9] The author has agreed to an amendment that cross-references the existing definition for "social media platform, to replace the reference in the bill to "social network."

Carved out of this obligation is digital content that a candidate for elective office posts or shares that portrays themself as doing or saying something that they did not do, so long as there is a disclosure indicating that the content has been manipulated that meets certain specifications. However, not only is this content not subject to the requirement for a platform to block and prevent, but platforms are required to host such content and cannot prevent such material, with no exception for whether it violates the platforms terms and services. This provision similarly applies during the period starting 120 days before an election through election day.

b. Labeling materially deceptive content

Large online platforms are also required to develop and implement procedures for labeling materially deceptive content as inauthentic, fake, or false. This applies to such content that meets the requirements from the above section but falls outside of the specified time range or that does not meet the requirements but that appears within an advertisement or election communication. The person or entity who created the materially deceptive content must have also done so knowing it was false or with reckless disregard for the truth. The rebuttable presumption again applies. And, for this obligation to trigger, the large online platform must have known or should have known that the materially deceptive content meets these requirements.

The label must allow users to click or tap on it and to inspect all available provenance data about the content in an easy-to-understand format. The labeling requirement applies during specified time periods: (1) the period starting one year before the election through election day; (2) that period through the 60th day after the election, if it depicts or pertains to elections officials, the electoral college process, a voting machine, ballot, voting site, or other property or equipment related to an election, or the canvass of the vote; and (3) during a governmental process related to redistricting, as provided.

c. Retention requirement

Content that either requires such a label or that must be blocked and prevented from being posted and shared must be retained by the platform for not less than five years from the relevant election. Platforms must share the content, upon request, with the Secretary of State, the Fair Political Practices Commission, and researchers.

d. Reporting mechanism

Lastly, the bill requires a large online platform to provide an easily accessible way for California residents to report to that platform content subject to the above provisions that was not blocked or labeled as required. The online platform shall respond to the person who made the report, within 36 hours, describing any action taken or not taken by the online platform.

e. Enforcement

The bill provides standing to candidates, elected officials, or elections officials who have made reports but who have either not received a timely response or who disagree with it to bring an action for injunctive and other equitable relief. The Attorney General, district attorneys, and city attorneys are also so authorized. A prevailing plaintiff is entitled to attorneys' fees and costs. Such actions are given precedence in the courts.

However, plaintiffs in such actions are required to establish a violation by clear and convincing evidence.

3. Legal concerns

Concerns have been raised about whether the bill runs afoul of federal statutory and constitutional law. Namely, whether the bill is preempted by Section 230 of the Communications Decency Act, 47 U.S.C. § 230 and the First Amendment to the United States Constitution.

a. Section 230

Section 230 does not apply to the *users* of social media (or the internet generally), but rather applies to the *platforms themselves*. In the early 1990s, prior to the enactment of Section 230, two trial court orders—one in the United States District Court for the Southern District of New York, and New York state court—suggested that internet platforms could be held liable for allegedly defamatory statements made by the platforms' users if the platforms engaged in any sort of content moderation (e.g., filtering out offensive material).[10] In response, two federal legislators and members of the burgeoning internet industry crafted a law that would give internet platforms immunity from liability for users' statements, even if they might have reason to know that statements might be false, defamatory, or otherwise actionable.[11] The result—Section 230—was relatively uncontroversial at the time, in part because of the relative novelty of the internet and in part because Section 230 was incorporated into a much more controversial internet regulation scheme that was the subject of greater debate.[12]

---

[10] *See Cubby, Inc. v. Compuserve, Inc.* (S.D.N.Y. 1991) 776 F.Supp. 135, 141; *Stratton Oakmont v. Prodigy Servs. Co.* (N.Y. Sup. Ct., May 26, 1995) 1995 N.Y. Misc. LEXIS 229, *10-14. These opinions relied on case law developed in the context of other media, such as whether bookstores and libraries could be held liable for distributing defamatory material when they had no reason to know the material was defamatory. (*See Cubby, Inc.*, 776 F. Supp. at p. 139; *Smith v. California* (1959) 361 U.S. 147, 152-153.)

[11] Kosseff, The Twenty-Six Words That Created The Internet (2019) pp. 57-65.

[12] *Id.* at pp. 68-73. Section 230 was added to the Communications Decency Act of 1996 (title 5 of the Telecommunications Act of 1996, Pub. L. 104-104, 110 Stat. 56), which would have imposed criminal liability on internet platforms if they did not take steps to prevent minors from obtaining "obscene or indecent" material online. The Supreme Court invalidated the CDA, except for Section 230, on the basis that it violated the First Amendment. (*See Reno v. ACLU* (1997) 521 U.S. 844, 874.)

The crux of Section 230 is laid out in two parts. The first provides that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[13] The second provides a safe harbor for content moderation, by stating that no provider or user shall be held liable because of good-faith efforts to restrict access to material that is "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected."[14]

Together, these two provisions give platforms immunity from any civil or criminal liability that could be incurred by user statements, while explicitly authorizing platforms to engage in their own content moderation without risking that immunity. Section 230 specifies that "[n]o cause of action may be brought and no liability may be imposed under any State law that is inconsistent with this section."[15] Courts have applied Section 230 in a vast range of cases to immunize internet platforms from "virtually all suits arising from third-party content."[16]

This bill provides for the potential liability of platforms for failing to block and prevent certain content from being posted or shared by users. If a user's content qualifies as materially deceptive, and other conditions are met, then the platform can be held liable for it.

Supporters point to the fact that monetary damages are not available and injunctive relief is essentially the only remedy available. The bill does allow for attorneys' fees and costs, which could be considered the type of liability that triggers Section 230's preemptive effect. The author has agreed to amendments that remove these remedies, leaving only injunctive relief. While courts, including the California Supreme Court, have found Section 230 immunity can extend to liability for solely injunctive relief, it is far from settled law in the country.[17]

In addition, the bill provide that if the platform engages in content moderation that restricts access to a candidate's deceptive portrayal of themselves (with the required disclosure and during the applicable time period), the platform can be held liable for that content moderation decision, regardless of the justification. As discussed below, the author has agreed to an amendment that removes this provision.

Ultimately, the bill is likely to face challenge on these grounds but these amendments work toward insulating the bill from such a challenge.

---

[13] *Id.*, § 230(c)(1).
[14] *Id.*, § 230(c)(1) & (2).
[15] *Id.*, § 230(e)(1) & (3).
[16] Kosseff, *supra*, fn. 13, at pp. 94-95; *see, e.g.*, *Doe v. MySpace Inc.* (5th Cir. 2008) 528 F.3d 413, 421-422; *Carfano v. Metrospalsh.com, Inc.* (9th Cir. 2003) 339 F.3d 1119, 1125; *Zeran v. America Online, Inc.* (4th Cir. 1997) 129 F.3d 327, 333-334.
[17] *Hassell v. Bird* (2018) 5 Cal. 5th 522, 547.

b. First Amendment

The First Amendment, as applied to the states through the Fourteenth Amendment, prohibits Congress or the states from passing any law "abridging the freedom of speech."[18] "[A]s a general matter, the First Amendment means that government has no power to restrict expression because of its message, its ideas, its subject matter, or its content."[19] However, while the amendment is written in absolute terms, the courts have created a handful of narrow exceptions to the First Amendment's protections, including "true threats,"[20] "fighting words,"[21] incitement to imminent lawless action,[22] defamation,[23] and obscenity.[24] Moreover, the First Amendment not only protects the right to speak, as a logical corollary it protects the "right to receive information and ideas."[25] Expression on the internet is given the same measure of protection granted to in-person speech or statements published in a physical medium.[26]

"Laws that burden political speech are subject to strict scrutiny, which requires the Government to prove that the restriction furthers a compelling interest and is narrowly tailored to achieve that interest."[27] Content-based restrictions subject to strict scrutiny are "presumptively unconstitutional."[28] California courts have been clear that political expression in the context of campaigns of any manner should be given wide latitude:

> Hyperbole, distortion, invective, and tirades are as much a part of American politics as kissing babies and distributing bumper stickers and pot holders. Political mischief has been part of the American political scene since, at least, 1800.
>
> In any election, public calumny of candidates is all too common. "Once an individual decides to enter the political wars, he subjects himself to this kind of treatment. . . . [D]eeply ingrained in our political history is a tradition of free-wheeling, irresponsible, bare knuckled, Pier 6, political brawls." To endure the animadversion, brickbats and skullduggery of a given campaign, a politician must be possessed with the skin of a

---

[18] U.S. Const., 1st & 14th amends.
[19] *Ashcroft v. American Civil Liberties Union* (2002) 535 U.S. 564, 573.
[20] *Snyder v. Phelps* (2011) 562 U.S. 443, 452.
[21] *Cohen v. California* (1971) 403 U.S. 15, 20.
[22] *Virginia v. Black* (2003) 538 U.S. 343, 359.
[23] *R.A.V. v. St. Paul* (1992) 505 U.S. 377, 383.
[24] *Ibid.*
[25] *Stanley v Georgia* (1969) 394 U.S. 557, 564. Internal citations omitted
[26] *Reno v. ACLU* (1997) 521 U.S. 844, 870.
[27] *Citizens United v. FEC* (2010) 558 U.S. 310, 340. Internal citations omitted. It should be noted that while not controversial for the principle cited herein, this opinion is widely criticized for further tilting political influence toward wealthy donors and corporations.
[28] *Reed v. Town of Gilbert* (2015) 135 S.Ct. 2218, 2226 (*Reed*).

> rhinoceros. Harry Truman cautioned would-be solons with sage advice about the heat in the kitchen.
>
> Nevertheless, political campaigns are one of the most exhilarating phenomena of our democracy. They bring out the best and the worst in us. They allow candidates and their supporters to express the most noble and, lamentably, the most vile sentiments. They can be fractious and unruly, but what they yield is invaluable: an opportunity to criticize and comment upon government and the issues of the day.
>
> The candidate who finds himself or herself the victim of misconduct is not without a remedy. Those campaign tactics which go beyond the pale are sanctionable under FPPC laws.
>
> It is abhorrent that many political campaigns are mean-spirited affairs that shower the voters with invective instead of insight. The elimination from political campaigns of opprobrium, deception and exaggeration would shed more light on the substantive issues, resulting in a more informed electorate. It would encourage more able people to seek public office. But to ensure the preservation of a citizen's right of free expression, we must allow wide latitude.[29]

The United States Supreme Court has emphasized the extraordinary protection afforded to political speech:

> Discussion of public issues and debate on the qualifications of candidates are integral to the operation of the system of government established by our Constitution. The First Amendment affords the broadest protection to such political expression in order "to assure [the] unfettered interchange of ideas for the bringing about of political and social changes desired by the people." Although First Amendment protections are not confined to "the exposition of ideas," "there is practically universal agreement that a major purpose of that Amendment was to protect the free discussion of governmental affairs,… of course includ[ing] discussions of candidates…." This no more than reflects our "profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open." In a republic where the people are sovereign, the ability of the citizenry to make informed choices among candidates for office is essential, for the identities of those who are elected will inevitably shape the course that we follow as a nation. As the Court observed in *Monitor Patriot Co. v. Roy*, 401 U.S. 265, 272 (1971), "it can hardly be doubted that the constitutional guarantee has its fullest and

[29] *Beilenson v. Superior Court* (1996) 44 Cal. App. 4th 944, 954-55. Internal citations omitted.

> most urgent application precisely to the conduct of campaigns for political office."[30]

This protection does not end where the truth of the speech does. "Although false statements of fact, by themselves, have no constitutional value, constitutional protection is not withheld from all such statements."[31] For instance, in the seminal opinion in *New York Times Co. v. Sullivan* (1964) 376 U.S. 254, 279-80, the court found the Constitution requires a rule that "prohibits a public official from recovering damages for a defamatory falsehood relating to his official conduct unless he proves that the statement was made with 'actual malice' -- that is, with knowledge that it was false or with reckless disregard of whether it was false or not. The Supreme Court has expounded on this principle, providing nuance based on the knowledge of the speaker:

> Truth may not be the subject of either civil or criminal sanctions where discussion of public affairs is concerned. And since ". . . erroneous statement is inevitable in free debate, and . . . it must be protected if the freedoms of expression are to have the 'breathing space' that they 'need . . . to survive' . . . ," only those false statements made with the high degree of awareness of their probable falsity demanded by *New York Times* may be the subject of either civil or criminal sanctions. For speech concerning public affairs is more than self-expression; it is the essence of self-government. The First and Fourteenth Amendments embody our "profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open, and that it may well include vehement, caustic, and sometimes unpleasantly sharp attacks on government and public officials."
>
> The use of calculated falsehood, however, would put a different cast on the constitutional question. Although honest utterance, even if inaccurate, may further the fruitful exercise of the right of free speech, it does not follow that the lie, knowingly and deliberately published about a public official, should enjoy a like immunity. At the time the First Amendment was adopted, as today, there were those unscrupulous enough and skillful enough to use the deliberate or reckless falsehood as an effective political tool to unseat the public servant or even topple an administration. That speech is used as a tool for political ends does not automatically bring it under the protective mantle of the Constitution. For the use of the known lie as a tool is at once at odds with the premises of democratic government and with the orderly manner in which economic, social, or political change is to be effected. Calculated falsehood falls into that class of utterances which "are no essential part of any exposition of ideas, and are

[30] *Buckley v. Valeo* (1976) 424 U.S. 1, 14-15. Internal citations omitted.
[31] *People v. Stanistreet* (2002) 29 Cal. 4th 497, 505.

> of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality. . . .” Hence the knowingly false statement and the false statement made with reckless disregard of the truth, do not enjoy constitutional protection.[32]

As stated, a restriction can survive strict scrutiny only if it uses the least-restrictive means available to achieve a compelling government purpose.[33] This bill implicates both the right to speak about elections, as well as the right to receive information regarding them. The bill is aimed at protecting the integrity of our elections, arguably a clearly compelling governmental interest. The question is whether the bill sufficiently tailors its provisions to effectuating that goal.

The bill seeks to prevent “materially deceptive content,” which is audio or visual media that is digitally created or modified, and that includes, but is not limited to, deepfakes and chatbots, such that it would falsely appear to a reasonable person to be an authentic record of the content depicted in the media, when it portrays candidates or elections officials doing or saying something they did not do or say. However, it does not target the person creating, posting, or sharing such content, but the platforms that host it. The bill attempts to tailor itself to the boundaries sketched out above. For instance, it imposes liability on the platforms only where they knew or should have known the content qualified as “materially deceptive content.” However, this falls short of the malice standard set forth in *Sullivan*, establishing something akin to a negligence standard instead.

The bill does impose a malice requirement but on the person or entity who created the content, requiring that they created it knowing it was false or with reckless disregard for the truth. However, liability is not imposed on the creator, nor even the one posting or sharing the content, but the social media platform allowing it on their platform. Further undercutting this element, there is a rebuttable presumption that the person who created it acted with malice if the content causes “a reasonable person to have a fundamentally different understanding or impression of the content than the person would have if hearing or seeing an authentic version of the content.” Therefore, the bill puts the onus on the platform to establish that the creator of the content, a person or entity the platform may not even have a relationship with or know, did not act with malice. Many of the relevant cases stress that the level of burden placed on a defendant to defend their political speech is a factor to consider. For instance, the following was stated in *Sullivan*:

> A rule compelling the critic of official conduct to guarantee the truth of all his factual assertions -- and to do so on pain of libel judgments virtually

---

[32] *Garrison v. Louisiana* (1964) 379 U.S. 64, 74-75. Internal citations omitted.
[33] *United States v. Playboy Entertainment Group* (2000) 529 U.S. 803, 813.

> unlimited in amount -- leads to a comparable "self-censorship." Allowance of the defense of truth, with the burden of proving it on the defendant, does not mean that only false speech will be deterred. Even courts accepting this defense as an adequate safeguard have recognized the difficulties of adducing legal proofs that the alleged libel was true in all its factual particulars.[34]

While the plaintiff is required to prove their case by clear and convincing evidence, the standards above place a burden on platforms to establish facts potentially well outside their bounds of knowing.

The California Initiative for Technology & Democracy (CITED), the sponsor of the bill, argues the case:

> AB 2655's approach is narrowly tailored and does not extend the law to hot button controversies or inflammatory claims – it does not ask social media platforms to adjudicate controversial opinions post by post. It simply stops the use of obviously, demonstrably untrue and provably false content meant to impermissibly influence our elections at peak election times. It is therefore respectful of the protections of the First Amendment and avoids concerns based on Section 230 of the Communications Decency Act.

Writing in opposition, ACLU California Action assesses the issue:

> Digitally modified content or content created using artificial intelligence (AI) tools is also entitled to [First Amendment] protections, unless the content falls within recognized First Amendment exceptions such as libel or fraud. The "novelty of deepfake technology and the speed with which it is improving" do not justify relaxing the stringent protections afforded to political speech by the First Amendment. The Supreme Court has held that "whatever the challenges of applying the Constitution to ever-advancing technology, 'the basic principles of freedom of speech and the press, like the First Amendment's command, do not vary' when a new and different medium for communication appears."
>
> The law has long made clear that the First Amendment was intended to create a wide berth for political speech because it is the core of our democracy. The First Amendment provides robust protection for speech of all kinds. Speech that is false, confusing, or which presents content that some find abhorrent, nevertheless maintains its constitutional protections as a driver of free discourse. This remains so no matter what the

[34] *N.Y. Times Co. v. Sullivan*, at 279.

> technology used to speak. Unfortunately, the provisions of AB 2655 as currently drafted threaten to intrude on those rights and deter that vital speech.

In response to these concerns, the author has agreed to amendments that remove the provision that applies this malice standard to the creators of the content and instead more closely hews the platform's basis for liability to the malice standard, holding the large online platform liable only if it knows that the materially deceptive content meets the requirements of the bill or acts with a reckless disregard for the truth.[35]

As the bill also requires platforms to allow certain potentially misleading content to be posted, the bill could be found to implicate the First Amendment rights of platforms in their editorial discretion. Two laws in Florida and Texas that similarly seek to prevent platforms from taking down certain content have been challenged. The consolidated case has been argued before the United States Supreme Court and an opinion is forthcoming. The 11th Circuit Court of Appeals laid out its assessment of the First Amendment implications of such laws:

> Social-media platforms like Facebook, Twitter, YouTube, and TikTok are private companies with First Amendment rights and when they (like other entities) "disclos[e]," "publish[]," or "disseminat[e]" information, they engage in "speech within the meaning of the First Amendment." More particularly, when a platform removes or deprioritizes a user or post, it makes a judgment about whether and to what extent it will publish information to its users—a judgment rooted in the platform's own views about the sorts of content and viewpoints that are valuable and appropriate for dissemination on its site. As the officials who sponsored and signed S.B. 7072 [the challenged Florida law] recognized when alleging that "Big Tech" companies harbor a "leftist" bias against "conservative" perspectives, the companies that operate social-media platforms express themselves (for better or worse) through their content-moderation decisions. When a platform selectively removes what it perceives to be incendiary political rhetoric, pornographic content, or public-health misinformation, it conveys a message and thereby engages in "speech" within the meaning of the First Amendment.
>
> Laws that restrict platforms' ability to speak through content moderation therefore trigger First Amendment scrutiny.[36]

---

[35] This amendment includes corresponding changes in the labeling section of the bill.
[36] *NetChoice, LLC v. AG, Fla.* (11th Cir. 2022) 34 F.4th 1196, 1210. Internal citations and quotations omitted.

As constitutional analysis is subject to changing norms and interpretations, especially in the more political charged federal judiciary of the day, it is inherently difficult to predict whether this law will be struck down for violating the protections of the First Amendment. However, it is safe to say it will likely face legal challenge and arguably be vulnerable thereto.

In order to insulate the bill from such challenge, the author has agreed to an amendment that simply provides that the bill does not apply to a candidate's portrayal of themselves doing or saying something that the candidate did not do or say, where it includes the required disclosure.

c. Additional concerns

The bill raises a few additional concerns. First, the bill requires platforms to retain all content they have prevented or blocked or labeled pursuant to the bill. This forced retention of information raises some thorny legal issues and may interfere with existing consumer rights. For instance, the CCPA, as amended by the CPRA, grants a series of rights to consumers, including the right to delete information held by businesses. In addition, given that the retention provision is essentially a government mandate on private businesses to seize certain information of private individuals, Fourth Amendment issues arguably arise. Furthermore, the bill requires platforms to hand over the content to specified government entities and even "researchers," upon request. There is no limitation that there be evidence of a crime or some other justification and no probable cause necessary to be provided the information. In response, the author has agreed to an amendment to remove this retention requirement.

In addition, it is unclear what exactly is required by the bill's requirement to block or prevent the *sending* of materially deceptive content. This could be read to apply to private messaging features of these platforms, essentially requiring platforms to scan private communications. This would raise serious privacy concerns. In response, the author has agreed to amendments that remove the "sending" element of the bill.

In addition, groups in opposition raise concerns that the bill presupposes a level of sophistication for technology that can detect AI-generated or manipulated content that simply does not exist. A coalition of industry associations, including NetChoice writes in opposition:

> AB 2655 appears to be based on the false assumption that online platforms definitively know whether any particular piece of content has been manipulated in such a way that is defined under the bill. While digital services may employ tools to identify and detect these materials with some degree of certainty, it is an evolving and imperfect science in its current form. AB 2655 also presumes that online platforms are an appropriate arbiter of deciding what constitutes accurate election

> information. However, most digital services are not equipped with the tools or expertise to make such judgments.

Oakland Privacy writes in opposition:

> The bill language offers that a technology company should be the judge, jury and executioner, although it may be unclear if the content is or is not generative AI-created and what role generative AI played in the content. It is unclear to us how any technology platform can be expected to know everything that every candidate in every city, county, state and federal election said and everywhere they went. Not to mention every other elected official in the state. If this is the basis for the removal of content by a technology platform, it is highly speculative and largely dependent on reports to the platform, which may be inaccurate, politically motivated, or malicious.
> We appreciate amendments to raise the bar for the knowledge level of online platforms. But we continue to have concerns on the other side of the spectrum: the removal of content that should not be removed and may well impact election results.
>
> In other words, the bill language is relying on two imprecise measures: technically scanning content for synthetic material with highly inaccurate tools, and real-life reports from the public, candidates and election officials and campaigns or chaos actors to power a broad censorship regime of blocking content. We cannot support that, even under the guise of defending democracy.

The opposition coalition also takes issue with the enforcement mechanism:

> [B]ecause AB 2655 is focused on enforcement against covered platforms and not the actors who are intentionally seeking to materially deceive other consumers, it is unlikely to meaningfully reduce the amount of election mis- and disinformation hosted online. While the June 11 amendments appear to attempt to address this issue, we do not believe the new language effectively resolves our concerns. For example, the bill now allows for a "rebuttable presumption" but still fails to effectively address and hold accountable the purveyors of deceptive content.

4. Support

CITED, the sponsor of the bill, writes:

> Those trying to influence campaigns – conspiracy theorists, foreign states, online trolls, and candidates themselves – are already creating and

> distributing election-threatening deepfake images, audio, and video content in the US and around the world. This threat is not imaginary: generative AI has been used in various ways – most of them deeply deceptive – to influence the national elections in Slovakia, Bangladesh, Argentina, Pakistan, and elsewhere, including in our own country. Examples of this occurring in U.S. elections include Ron Desantis using AI-generated images to attack his opponent in his presidential run, foreign states caught attempting to influence American politics through social media, and just this month, a supporter of former President Trump creating a deepfake image of Trump with Black Americans designed to persuade Black voters to support Trump.
>
> These examples demonstrate the power of generative AI-fueled disinformation to skew election results and weaken our faith in our democracy. We cannot let it undermine our elections here in California, and we are grateful you are leading the effort to try to stop it.
>
> AB 2655 strikes the right balance by seeking to ban, for a strictly limited time before and after elections, the online spread of the worst deepfakes and disinformation maliciously intended to prevent voters from voting or getting them to vote erroneously based on fraudulent content. The bill also requires that other fake online content related to elections and elections processes (such as redistricting), which is also designed to undermine election procedures and democratic institutions, must be labeled as fake, again just for a limited time. The bill only applies to the largest online platforms with the greatest reach of potential election disinformation, and we believe it is fully implementable today based on tools these companies already possess. The companies covered by the bill's requirements are all already subject to similar requirements under the European Union's Digital Services Act, which is designed to, among other things, crack down on election interference.

A coalition of groups in support, including the American Federation of State, County, and Municipal Employees (AFSCME) and NextGen CA, write:

> AB 2655 seeks to solve these problems by, for a limited time before and after elections, banning the online spread of the worst of the deepfakes and disinformation meant to prevent voters from voting or to deceive them based on fraudulent content, and requiring that other fake content to be labeled as such. The approach leans heavily on increasing transparency, with bans used at only the highest-leverage moments, making it narrowly tailored. Additionally, it does not extend the law to hot button controversies or inflammatory claims – just the depiction of demonstrably untrue and provably false content meant to impermissibly

> influence our elections, at peak times – and is therefore implementable and respectful of the protections of the First Amendment.

Writing in support, the Northern California Recycling Association explains the need for the bill:

> California is entering its first-ever generative Artificial Intelligence (AI) election, in which disinformation powered by generative AI would and will pollute our information ecosystems like never before. Deepfakes are a powerful and dangerous tool in the arsenal of those that want to wage disinformation campaigns, and they have the potential to wreak havoc on our democracy by attributing speech and conduct to a person that is false or that never happened. Advances in AI make it easy for practically anyone to generate this deceptive content, making it that much more important that we identify and restrict its spread before it has the chance to deceive voters and undermine our democracy.

SUPPORT

California Initiative on Technology and Democracy (sponsor)
AFSCME California
Asian Americans Advancing Justice - Asian Law Caucus
Asian Americans and Pacific Islanders for Civic Empowerment
Asian Law Alliance
Bay Rising
Board of Supervisors for the City and County of San Francisco
California Clean Money Campaign
California Environmental Voters
California State Sheriff's Association
California Voter Foundation
Center for Countering Digital Hate
Chinese Progressive Association
City and County of San Francisco Board of Supervisors
Courage California
Disability Rights California
Hmong Innovating Politics
Inland Empire United
League of Women Voters of California
Move (mobilize, Organize, Vote, Empower) the Valley
Nextgen California
Northern California Recycling Association
Partnership for the Advancement of New Americans
SEIU California
Techequity Action

Verified Voting
Young People's Alliance
Youth Power Project

OPPOSITION

ACLU California Action
California Chamber of Commerce
Computer & Communications Industry Association
Electronic Frontier Foundation
Internet Works
Netchoice
Oakland Privacy
Software & Information Industry Association
Technet

RELATED LEGISLATION

Pending Legislation:

SB 942 (Becker, 2024) establishes the California AI Transparency Act, requiring covered providers to create and make freely available an AI detection tool to detect content as AI-generated and to include disclosures in content generated by the provider's system. SB 942 is currently in the Assembly Judiciary Committee.

SB 970 (Ashby, 2024) ensures that media manipulated or generated by artificial intelligence (AI) technology is incorporated into the right of publicity law and criminal false impersonation statutes. The bill requires those providing access to such technology to provide a warning to consumers about liability for misuse. SB 970 was held on suspense in the Senate Appropriations Committee.

AB 2355 (Wendy Carrillo, 2024) requires committees that create, publish, or distribute a political advertisement that contains any image, audio, or video that is generated or substantially altered using artificial intelligence to include a disclosure in the advertisement disclosing that the content has been so altered. AB 2355 is currently in this Committee.

AB 2839 (Pellerin, 2024) prohibits a person, committee, or other entity from knowingly distributing an advertisement or other election communication that contains materially deceptive content, as defined and specified, with malice, except as provided, within 120 days of a California election, and in specified cases, 60 days thereafter. AB 2839 is currently in this Committee.

AB 2930 (Bauer-Kahan, 2024) requires, among other things, a deployer and a developer of an automated decision tool to perform an impact assessment for any automated

decision tool the deployer uses that includes, among other things, a statement of the purpose of the automated decision tool and its intended benefits, uses, and deployment contexts. AB 2930 requires a deployer to, at or before the time an automated decision tool is used to make a consequential decision, notify any natural person that is the subject of the consequential decision that an automated decision tool is being used to make, or be a substantial factor in making, the consequential decision and to provide that person with, among other things, a statement of the purpose of the automated decision tool. AB 2930 is currently in this Committee.

AB 3211 (Wicks, 2024) establishes the California Provenance, Authenticity and Watermarking Standards Act, which requires a generative AI system provider to take certain actions to assist in the disclosure of provenance data to mitigate harms caused by inauthentic content, including placing imperceptible and maximally indelible watermarks containing provenance data into content created by an AI system that the generative AI system provider makes available. AB 3211 also requires a large online platform, as defined, to, among other things, use labels to prominently disclose the provenance data found in watermarks or digital signatures in content distributed to users on its platforms, as specified. AB 3211 is currently in the Senate Appropriations Committee.

Prior Legislation: AB 730 (Berman, Ch. 493, Stats. 2019) prohibited the use of deepfakes depicting a candidate for office within 60 days of the election unless the deepfake is accompanied by a prominent notice that the content of the audio, video, or image has been manipulated. Additionally, AB 730 authorized a candidate who was falsely depicted in a deepfake to seek rapid injunctive relief against further publication and distribution of the deepfake.

PRIOR VOTES:

Senate Elections and Constitutional Amendments Committee (Ayes 6, Noes 1)
Assembly Floor (Ayes 56, Noes 1)
Assembly Appropriations Committee (Ayes 11, Noes 1)
Assembly Judiciary Committee (Ayes 9, Noes 0)
Assembly Elections Committee (Ayes 6, Noes 1)
****